

CRUI – Conferenza dei Rettori delle Università Italiane
Commissione Biblioteche di Ateneo
Gruppo di lavoro “Editoria elettronica”
Documento: “Lo stato dell'arte dell'editoria elettronica negli Atenei Italiani”
Contributo/REPORT di Nunzio Femminò: “Quadro tecnologico”

L'editoria elettronica è un settore di recente affermazione e ancora in corso di sviluppo dal punto di vista tecnologico. Accanto a prodotti proprietari sviluppati dai grandi editori del segmento STM (Science, Technology, Medicine), che per primi si sono lanciati in maniera massiccia nello sviluppo dell'editoria elettronica e dei servizi correlati, si stanno diffondendo i primi prodotti commerciali e liberi per la realizzazione di iniziative di editoria elettronica di dimensioni medio-piccole, sia autonome sia legate alla produzione intellettuale di enti quali le università.

Il quadro tecnologico attuale consente agli atenei di affiancare alla produzione editoriale tradizionale delle University Press (riviste cartacee, monografie e collane di studi, dispense e materiale didattico a stampa) anche una produzione digitale a costi meno elevati, con il vantaggio di poter ampliare il catalogo dei titoli offerti.

Accanto ai prodotti tradizionali che possono generare profitti commerciali, è infatti possibile oggi adottare uno dei numerosi nuovi applicativi di software libero open source e dare avvio a iniziative di editoria sostenibile, sia ad accesso aperto sia a basso prezzo (con l'integrazione di sistemi di print-on-demand), con il vantaggio di offrire servizi efficienti per l'utenza (ad esempio realizzando repositories per le dispense ed il materiale didattico) e di aumentare la visibilità ed il prestigio dell'ateneo (attraverso gli archivi aperti per la disseminazione degli articoli di ricerca e il software per la pubblicazione di riviste open access da parte dei propri gruppi di ricerca).

Com'è noto, e sarà sufficiente rifarsi al manifesto BOAI¹, le due strategie per l'attuazione dell'Open access sono rappresentate dall' *Open access self-archiving* e dall'*Open access Publishing*, la prima volta a promuovere il self-archiving, l'autoarchiviazione dei contributi di ricerca già sottoposti al controllo di qualità, la seconda, volta a favorire la creazione di nuove riviste ad accesso aperto e la conversione di quelle a pagamento già esistenti.

La prima strategia si esplica attraverso la creazione di archivi *aperti*, conosciuti meglio col termine “*Open Archive*”² che si avvalgono prevalentemente di

¹ BOAI, Budapest Open Access Initiative, (<http://www.soros.org/openaccess/>)

² « Che cosa è un archivio *e-prints*? E' un *open archive*, o *e-prints server*, ossia un archivio preposto al deposito dei documenti scientifici in forma elettronica, alla loro gestione e conservazione. Può essere usato anche per il deposito dei materiali didattici o delle tesi, anche se nasce come strumento per la disseminazione dei lavori di ricerca. Può essere organizzato a livello istituzionale o a livello disciplinare.», Antonella De Robbio, *Gli archivi e-print in Italia*, in *Bibliotime*, anno VII, 1, 2004.

software distribuiti con licenza GNU-GPL³ - semplici ed economici con installazioni standard - e sono usati per mettere gratuitamente a disposizione della rete gli e-prints⁴ depositati.

Le più Importanti caratteristiche di un *Open Archive, Istituzionale o Disciplinare*, sono le seguenti:

- *Self-Archiving*, un meccanismo di auto-sottomissione ma con la possibilità di attivare un efficiente sistema di peer-review.
- *Metadata standard*, MARC⁵ e Dublin-Core⁶ per condividere i contenuti anche ai più importanti motori di ricerca del web.
- *Il protocollo di comunicazione OAI-PMH*⁷ per permettere la interoperabilità e lo scambio di dati con altri Open Archive attraverso un sistema di harvesting.

L'organizzazione internazionale di riferimento per gli Open Archive è l'*OAI*⁸.

L'OAI nasce con l'intento di favorire l'interoperabilità degli archivi eterogenei e per facilitare la disseminazione degli *e-print*.

I principali attori previsti nell'architettura OAI sono due: da una parte ci sono i *Data Provider*, dall'altra troviamo i *Service Provider*. I primi (DP) rappresentano gli archivi presso i quali vengono depositati sia i *full text* dei lavori di ricerca (articoli, presentazioni a convegni, ecc) che i relativi metadata e sono, dunque, i contenitori dei dati (repository⁹); i secondi (SP) sono sistemi che si collocano logicamente ad un livello superiore rispetto ai DP in quanto

³ La GPL, General Public License (<http://www.gnu.org/copyleft/gpl.html>) è considerata da molti, se non tutti, la licenza basilare del software libero.

⁴ « Gli *e-prints* sono copie elettroniche di papers accademici, comprendono: i *pre-prints*, i *post-prints*, i *post-post-prints*. Gli *e-prints* posti dentro gli archivi di *e-prints* consentono agli autori di rendere le proprie produzioni di ricerca liberamente disponibili alla comunità scientifica internazionale, disseminandoli su scala mondiale, cosa impossibile per un lavoro su carta.», Antonella De Robbio, *Gli archivi e-print in Italia*, in *Bibliotime*, anno VII, 1, 2004.

⁵ MARC - Machine Readable Cataloging, è uno standard internazionale creato per facilitare lo scambio d'informazioni bibliografiche e la produzione di registrazioni bibliografiche grazie alla codificazione degli elementi bibliografici e alla struttura logica e fisica dell'insieme.

⁶ Dublin Core Metadata Element Set (<http://purl.org/dc>): standard di metadata (ancora in via di sviluppo e perfezionamento) composto da 15 elementi di base estesi anche a sottoelementi o qualificatori. Ciascun elemento è definito usando un set di 10 attributi ricavati dalla norma ISO 11179.

⁷ OAI-PMH (<http://www.openarchives.org/OAI/openarchivesprotocol.html>), the Open Archives Initiative Protocol for Metadata Harvesting: fornisce una cornice di interoperabilità indipendente dagli applicativi e basata sull'harvesting dei metadata.

⁸ OAI (<http://www.openarchives.org/>), Open Archives Initiative: nata e gestita da C.Lagoze (Cornell) e H.Van de Sompel (Los Alamos), finanziata a tratti da varie istituzioni, sviluppa e promuove gli standard per l'interoperabilità.

⁹ «A *repository* is a network accessible server that can process the 6 OAI-PMH requests in the manner described in this document. A repository is managed by a data provider to expose metadata to harvesters.», tratto dalla sezione "Definizioni e Concetti" della descrizione del protocollo

OAI

<http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm#DefinitionsConcepts>

offrono servizi, anche a valore aggiunto, tra i quali la raccolta e l'indicizzazione di metadata da altri Data Providers (harvesting).

L'harvester è dunque, un client che invia richieste secondo il protocollo OAI-PMH, che infatti, inoltra al repository una delle sei richieste previste dall'OAI-PMH (Identify, ListMetadataFormat, ListSets, GetRecord, ListIdentifiers, ListRecord) in formato HTML. Il repository può rispondere con diversi formati di metadata. Attraverso il comando ListMetadataFormat, per esempio, è possibile ricevere dal Data Provider la lista dei formati di metadata di esportazione disponibili. Attraverso il comando GetRecord, corredato degli opportuni parametri, è possibile ricevere localmente i records secondo il formato di esportazione richiesto o previsto in un file di testo, a sua volta "formattato" con il linguaggio XML, vettore comune in tutti gli harvester.

Nel caso in cui il formato di metadata di esportazione del repository interrogato coincide con il formato di gestione del proprio Data Provider, sarà possibile attivare il tool per l'upload del file TEXT/XML, ma se non coincide sarà necessario implementare uno script di configurazione per la conversione del formato di metadata.

In effetti, il processo di harvesting dei metadata è meno semplice e immediato di quanto sembri. Non avendo interfacce grafiche o tools a corredo, richiede, nella migliore delle ipotesi, alcuni passaggi da effettuare direttamente alla console con il probabile ausilio da parte di un tecnico informatico.

La seguente tabella riepilogativa delle principali caratteristiche e funzionalità dei software di gestione degli Archivi Aperti più diffusi è una rielaborazione di una più ampia panoramica disponibile sul sito della BOAI (http://www.soros.org/openaccess/software/OSI_Guide_to_Institutional_Repository_Software_v1/css/OSI_Guide_to_Institutional_Repository_Software_v1_10.htm).

	Archimede	ARNO	CDSware	Dspace	Eprints	Fedora
Home Page software		http://www.uba.uva.nl/arno	http://cdsware.cern.ch	http://dspace.org	http://software.eprints.org	http://www.fedora.info/
Developed by		Tiburg & Netherland's University	CERN – Geneve	MIT & HP	University of Southampton	Cornell & Virginia University
OAI-PMH version supported	OAI-PMH 2.0	OAI-PMH 2.0	OAI-PMH 2.0	OAI-PMH 2.0	OAI-PMH 2.0	OAI-PMH 2.0
39.50 protocol compliant	No	No	No	No	No	No
Open source license	GNU GPL	TBD	GNU GPL	BSD	GNU GPL	MPL
Latest version release date	May-04	Dec-03	Aug-02	Apr-04	Mar-02	Apr-04
Latest version number	1.0	1.0	0.0.9	1.2	2.3.6	1.2.1
Minimum hardware requirements	No specific requirements	No specific requirements	No specific requirements	No specific requirements	No specific requirements	No specific requirements
SAN support	No	No	No	Yes	Yes	Yes
Operating system (tested)	Linux/ Windows	Linux/ Solaris	Linux/ Solaris	UNIX/ MacOSX/ Windows	GNU/ Linux/ Solaris	Unix/ MacOSX/ Windows
Programming language	Java	Perl	Python/PHP	Java	Perl	Java
Database	Many	Oracle 8i	MySQL	PostgreSQL/ Oracle3	MySQL	MySQL/ McKoi/ Oracle

Web server	Any	Apache	Apache/PHP, Python	Any	Apache 1.3.2	Tomcat 4.1
Java servlet engine	Any	N/A	N/A	Any	N/A	Tomcat 4.1
Search engine	Lucene	N/A	cdsware	Lucene	N/A	Database
Other	Lius, OAICat, Torque, Struts	N/A	WML: Website META Language	OAICat, SRW	N/A	N/A
Clients supported	Any browser with minimal CSS & Javascript support	Any browser with minimal CSS & Javascript support	All HTML 4.0 clients	All web browsers	Netscape, Mozilla, IE, Lynx	Web browsers and SOAP clients
Number of installations	1	7	7+	20+	140	20
Geographic coverage	Canada	Netherlands	Europe & US	Worldwide	Worldwide	Worldwide
Set-up/Installation						
Automated installation script	Yes	Yes	Yes	Yes	Yes	Yes
Password administration	Yes	Yes	Yes	Yes	Yes	Yes
System-assigned passwords	Yes	No	Yes	Yes	No	No
User selected passwords	Yes	Yes	Yes	Yes	Yes	Yes
User registration verification/Other security mechanisms	Database table	LDAP and/or ARNO registry	MySQL table/ Apache ACL	email/ X.509	MySQL table	No
Submission Support						
Edit user profile	Yes	No	Yes	Yes	Yes	No
Submission Stages	Pending, Approved		Submit, Modify, Revise, Approve, etc.	Assemble, Pending, Approved		Ingest, Create, Modify, Activate, Deactivate
Segregated submission workspace	Yes	Yes	Yes	Yes	Yes	Yes
Submission roles	Administrator, Community Administrators, User	Contributors, Editors, Administrators, Site Managers	Submitters, Moderators, Reviewers, Approvers, Administrators	Submitters, Reviewers, Approvers, Editors	User, Editor, Administrator	Administrator
Email notification for submitters	Yes	Only during registration	Yes	Yes	Yes	No
Email notification for content administrators	Yes	Yes	Yes	Yes	Yes	No
Personalized system access for registered users	Yes	Yes	Yes	Yes	Yes	No
View pending content submissions	Yes	Yes	Yes	Yes	Yes	No
View approved content	Yes	Yes	Yes	Yes	Yes	No
View pending content administration tasks	Yes	Yes	Yes	Yes		No
System-generated usage statistics	No	Yes	No	Yes	No	Yes
Usage reports	No	No	No	Yes	No	No
File formats ingested	All	All	All	All	All	All
Submitted items can comprise multiple files	Yes	Yes	Yes	Yes	Yes	Yes
Metadata						
Metadata schema supported	Qualified Dublin Core	Dublin Core	Standard Marc21	Qualified Dublin Core	Dublin Core	Dublin Core
Support for extended metadata	No	Yes	Yes	Custom	Yes	Yes
Metadata review support	Yes	Yes	Yes	Yes	Accept, Edit, Bounce (require	No

					changes), Delete	
Metadata export	Yes	Yes	OAI-Marc export	METS & Custom XML Schema	Custom XML Schema	Yes
Disallow metadata harvesting	Yes	Yes	Yes	Yes	Yes	Yes
Add/delete metadata fields	Yes	Yes	Yes	Yes	Yes	Yes
Set default values for metadata	No	Yes	Yes	Yes		No
Supports Unicode character set for metadata	Yes	Partial	Yes	Yes	Yes	Yes
Modify interface "look & feel"	Yes	Yes	Yes	Yes	Yes	Yes
Supports multiple language interfaces	Yes	No	Yes	Yes	Yes	Yes
End user document folders	Yes	No	Yes	No	No	No
Discussion forum support	No	No	No	No	Yes	No
Search Capability						
Full text	Yes	No	Yes	Yes	Yes	No
Boolean logic	Yes	No	Yes	No	No	No
Search all descriptive metadata	Yes	No	Yes	Yes	Yes	Yes
Search selected metadata fields	Yes	No	Yes	Yes	Yes	Yes
Indexed by Google/ Other Search Engines	Possible	Possible	Possible	Yes	Possible	Possible
System-assigned identifiers	Yes	Yes	Yes	Yes	Yes	Yes
Data integrity checks	No	No	No	MD5 checksum	MD5 checksum	SIP schema validation
Documentation/Support						
Documentation/manual	Yes	Yes	Yes	Yes	Yes	Yes
Listserv	Yes	Yes	Yes	Yes	Yes	Yes
Bug track/ feature request system	Yes	No	Yes	Yes	No	Yes
Formal support/help desk	No	No	For fee	No	No	Yes

Alla luce del confronto su esposto, ho ritenuto di sottolineare alcuni aspetti. In verde, ho evidenziato la versione 2 del protocollo OAI-PMH comune a tutti i software, il rilascio della licenza d'uso GNU-GPL presente nella maggior parte dei software e i data-base di riferimento che in tutti i repository è di natura "free".

In giallo sono indicate le installazioni italiane.

Le cratteristiche evidenziate in rosso sono, invece, quelle che, a mio avviso, rappresentano l'unica nota "stonata". Infatti si riscontra che lo standard di gestione dei metadata e quello di esportazione non sono uniformi tra i vari software.

La portata del mio rilievo non è grave ma ritengo che l'attivazione semplice e immediata delle operazioni di harvesting, sarebbe più efficientemente garantita da una "lingua comune" tra data provider e service provider.

La seconda strategia della BOAI, ossia la pubblicazione in riviste ad accesso aperto, può essere realizzata sia incentivando gli autori a pubblicare su riviste Open Access già esistenti¹⁰, sia supportando iniziative editoriali autonome dei gruppi di ricerca italiani, fornendo loro l'infrastruttura tecnologica che consente la pubblicazione a basso costo.

Oltre al progetto DAFNE¹¹, sviluppato da alcuni atenei italiani ma il cui software non è ancora in distribuzione, una carrellata del software libero disponibile per questo tipo di iniziative editoriali è disponibile sul sito di SPARC¹². Per questi strumenti non esiste al momento uno studio comparativo paragonabile a quello commissionato dall'OSI per i repositories.

Un'altra fonte di informazioni sulle risorse disponibili in rete, che offre un quadro completo in termini di software, tools, documentazione di supporto e promozione dell'Open Access, link a progetti italiani ed esteri, è disponibile sul sito di AEPIC¹³, progetto nato nell'ambito del CILEA ed arricchito dalla collaborazione del CASPUR, che ha consentito la realizzazione del primo service provider per la comunità accademica italiana, e la piattaforma PLEIADI (che verrà presentata al convegno di Messina il 4-5 novembre prossimi). Sul sito di AEPIC sono inoltre illustrati diversi servizi per la progettazione e l'implementazione di servizi di editoria elettronica per gli atenei.

¹⁰ Un elenco aggiornato è disponibile sul sito <http://www.doaj.org>

¹¹ Il progetto DAFNE – District Architecture For Networked Editions (<http://dafne.cab.unipd.it/>) è la sintesi di tre complessi progetti elaborati rispettivamente dalle Università di Padova e di Firenze, dall'Università di Bologna e dalla Biblioteca Centrale di Firenze che, partendo da punti di vista simili ma distinti, tendevano a realizzare alcuni significativi segmenti di un articolato prodotto/processo che in letteratura viene definito come Biblioteca Digitale Nazionale

¹² SPARC - the Scholarly Publishing and Academic Resources Coalition (Publishing Resources, <http://www.arl.org/sparc/resources/pubres.html>)

¹³ AEPIC - Academic E-Publishing Infrastructures (<http://www.aepic.it/>)